

(12) UK Patent Application (19) GB (11) 2 316 575 (13) A

(43) Date of A Publication 25.02.1998

(21) Application No 9717929.5

(22) Date of Filing 22.08.1997

(30) Priority Data

(31) 08239938 (32) 23.08.1996 (33) JP

(71) Applicant(s)

Kokusai Denshin Denwa Kabushiki Kaisha

(Incorporated in Japan)

3-2 2-chome Nishishinjuku, Shinjuku-ku,
Tokyo-to 163-03, Japan

(72) Inventor(s)

Shingo Kuroiwa
Masaki Naito
Seiichi Yamamoto
Shin-ichi Sakayori
Masanobu Fujioka(51) INT CL^B

H04M 1/65

(52) UK CL (Edition P)

H4K KBHX
G4R REX R1F

(56) Documents Cited

GB 2260670 A GB 2184412 A US 5479490 A
US 5274695 A

(58) Field of Search

UK CL (Edition O) H4K KBHE KBHF KBHX KF42 KF50A
KF50C KF50X KOE
INT CL^B: H04M
ONLINE: WPI

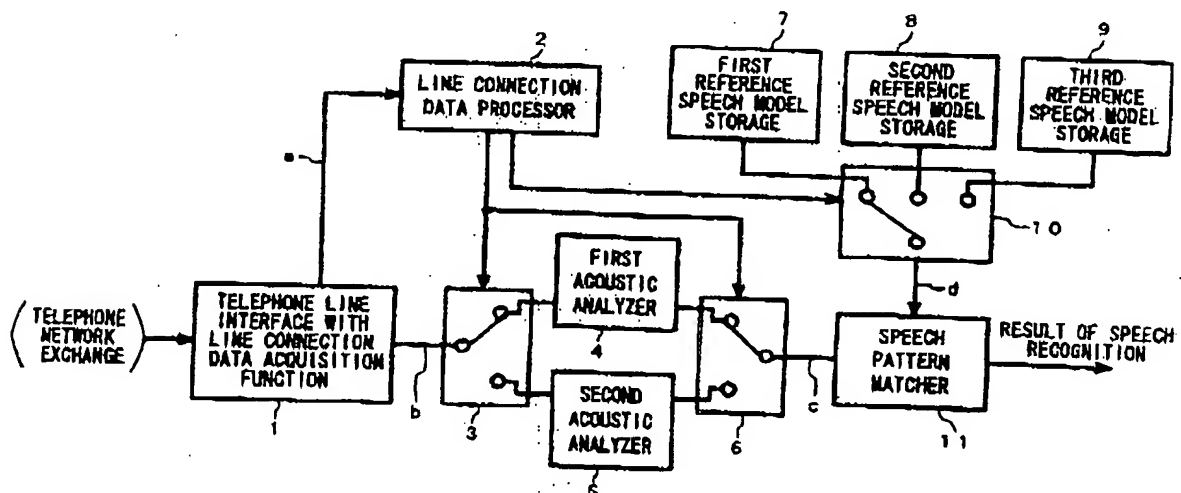
(74) Agent and/or Address for Service

Boulton Wade Tennant
27 Fumival Street, LONDON, EC4A 1PQ,
United Kingdom**(54) Telephone speech recognition**

(57) A telephone speech recognition system is provided which has a high level of speech recognition regardless of various conditions of a telephone line.

The system comprises speech analyzers 4 and 5, and reference speech model storages 7 to 9 corresponding to line connection data. A telephone line interface 1 having a line connection data acquisition function analyzes a call received from the telephone line for identifying the country, route, and other information of the call and transmits those line connection data to a line connection data processor 2. The line connection data processor 2 selects one of the acoustic analyzers 4 and 5 in response to the line connection data from the interface 1 and also one of the speech model storages 7 to 9. A speech pattern matcher 11 compares an acoustic vector train output of the selected acoustic analyzer with the speech models given from the selected reference speech model storage for speech recognition.

FIG. 1



Best Available Copy

GB 2 316 575 A

FIG. 1

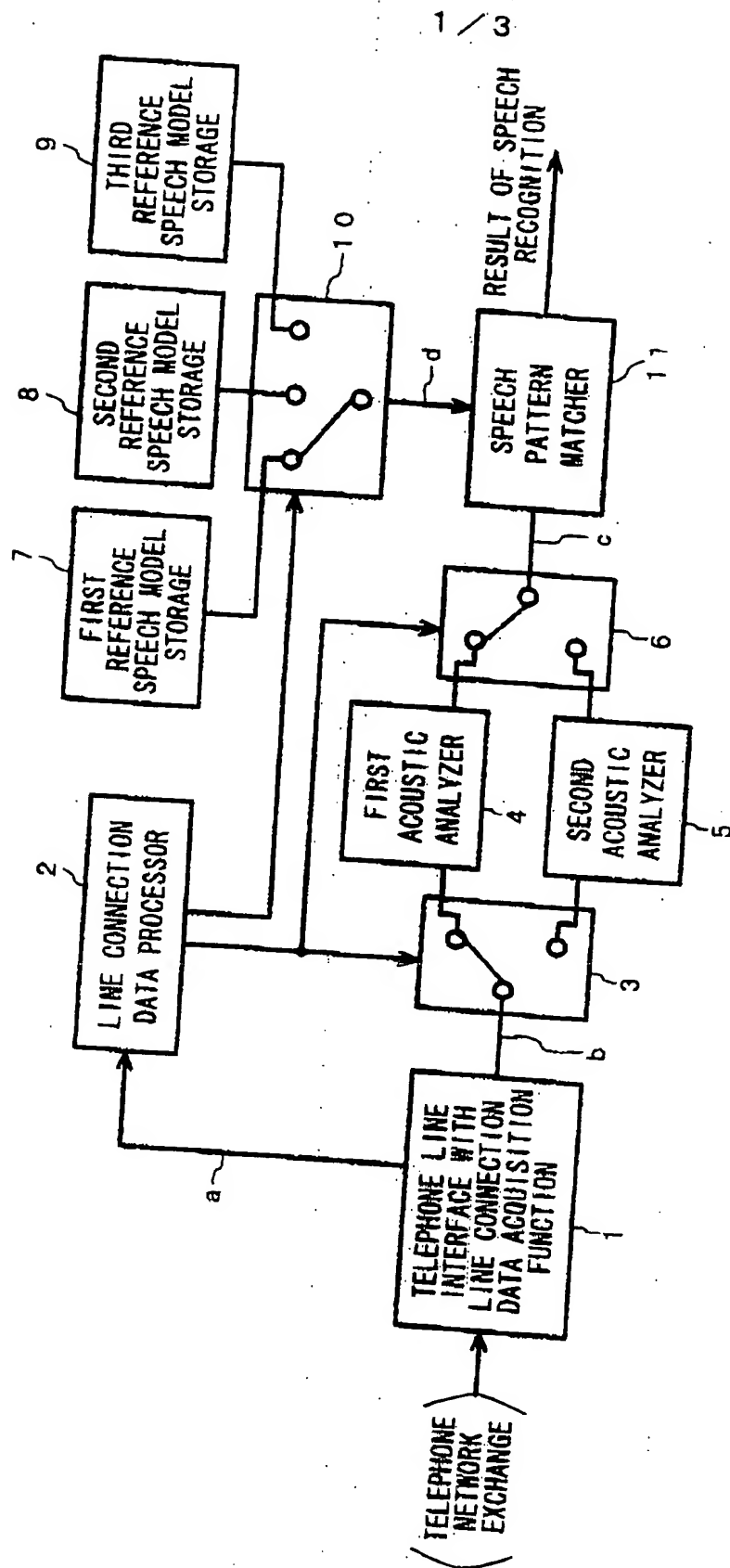
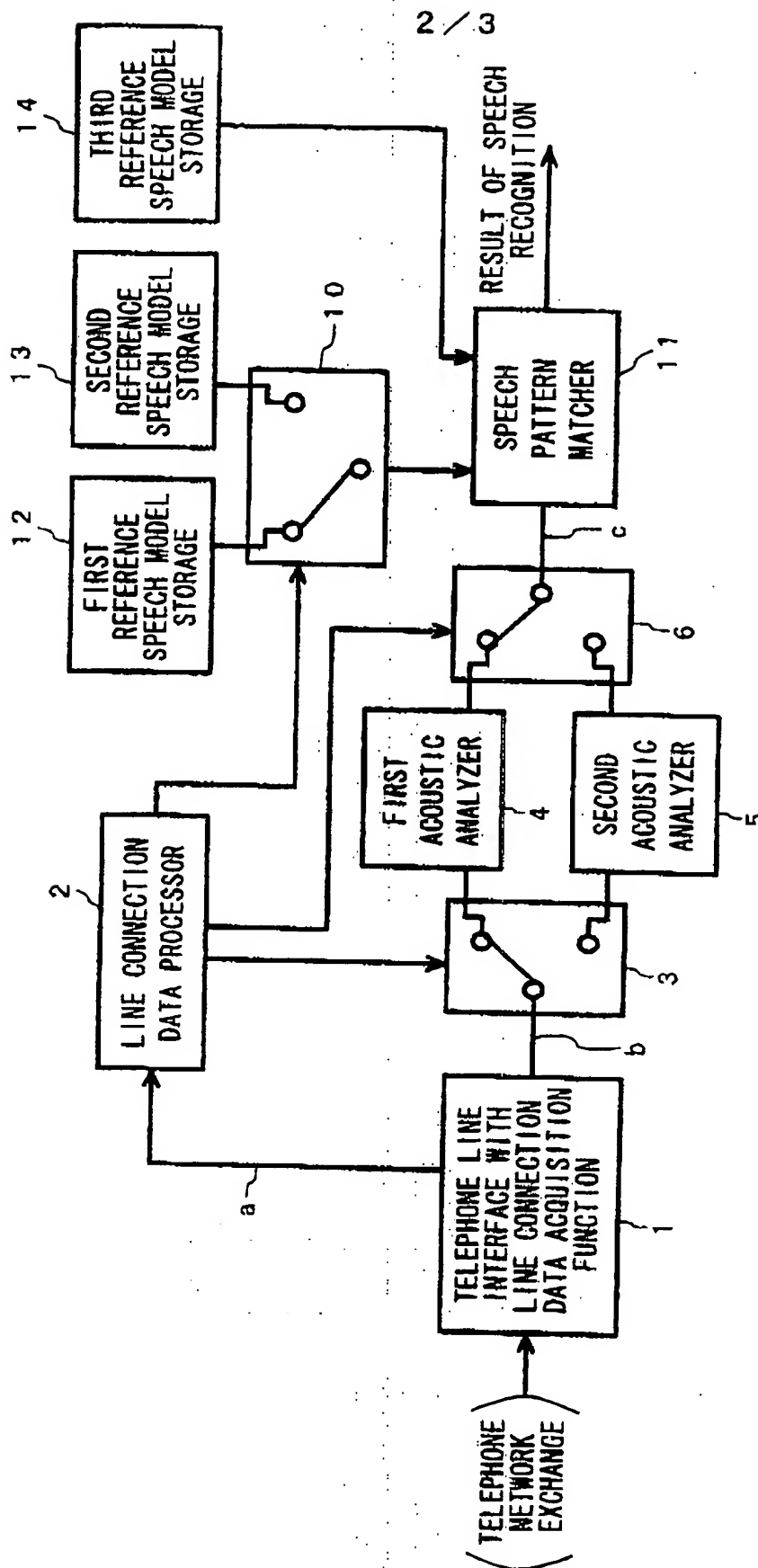


FIG. 2



2316575**TELEPHONE SPEECH RECOGNITION SYSTEM**

The present invention relates to a telephone speech recognition system and particularly, a telephone speech recognition system for use with a telephone line, network, or switchboard.

A conventional speech recognition system connected to a telephone line is now explained referring to Fig. 3. When the speech of a caller is received from a telephone line and has to be identified, it is transmitted via a telephone line interface 31 to the speech recognition system. The conventional speech recognition system substantially comprises an acoustic analyzer 32, a speech pattern matcher 33, and a reference speech model storage 34.

The speech of the caller introduced from the telephone line interface 31 is first fed to the acoustic analyzer 32. In the acoustic analyzer 32, the speech is divided at equal intervals of substantially 10 milliseconds (ms) on the basis of a Hamming window of about 25 ms and subjected to acoustic analysis such as cepstrum analysis to produce an acoustic vector train which is then supplied to the speech pattern matcher 33. The reference speech model storage 34 saves speech models such as HMMs (hidden Markov models). The speech pattern matcher 33 collates the acoustic vector train with the speech models saved in the reference speech model storage 34. Consequently, a succession of symbols representing the speech models at highest likelihood are released as outputs of the speech recognition.

In the conventional manner, the speech recognition is however carried out regardless of the conditions of line connection from a caller at

the other end of the line to the telephone line interface 31 at the entrance of the speech recognition system. More specifically, the collating action of the speech pattern matcher 33 with the speech models saved in the reference speech model storage 34 is executed without concerning any of the conditions of line connection. Accordingly, when particular noises from the telephone line are involved or there is a difference in the frequency characteristic between the conventional telephone line and a line used with a system such as a mobile telephone set, a desired level of speech recognition will hardly be accomplished.

It is particularly difficult for such a conventional speech recognition system to recognize voice data in a call received from the international telephone line which varies depending on terminal and line systems of a country.

It is therefore desirable to provide a telephone speech recognition system capable of performing a high level of speech recognition without being affected by various conditions of a telephone line while eliminating the foregoing drawbacks of the prior art.

According to the present invention there is provided a telephone speech recognition system for recognizing speech data received from a telephone line comprising: a telephone line interface connected to the telephone line for detecting line connection data; a plurality of acoustic analyzers having means for removing noise derived from the line characteristics and/or the route characteristics; and a line connection data processor responsive to the line connection data from the telephone line interface for selecting the acoustic analyzers. In addition, a plurality of reference speech model storages for saving speech models

corresponding to the line characteristics and/or the route characteristics may be provided.

A telephone speech recognition system embodying the present invention may allow the telephone line connection data to be used for selecting a desired acoustic analyzer and a pertinent speech model, thus enhancing the quality of speech recognition.

Embodiments of the invention will now be described by way of example with reference to the drawings, in which:

Fig. 1 is a block diagram showing a first embodiment of the present invention;

Fig. 2 is a block diagram showing a second embodiment of the present invention; and

Fig. 3 is a block diagram showing a conventional speech recognition system.

Embodiments of the invention will be described in detail with reference to the accompanying drawings.

Fig. 1 is a block diagram showing one embodiment of the present invention.

As shown, a line interface 1 having a known function of line connection data acquisition is connected to a telephone line, a network or a switching board. The telephone line interface 1 with the line connection data acquisition function examines line connection data of a received call including the telephone number of a caller, the interconnection in a private branch exchange (PBX), and (when the call is an international call,) the country of the caller and the route of

transmission (e.g. via satellite links or underwater cable links). The line connection data a extracted by the telephone line interface 1 with line connection data acquisition function is transmitted to a line connection data processor 2 while a speech data b of the call is given to a first switching unit 3.

In response to the line connection data a from the telephone line interface 1 with line connection data acquisition function, the line connection data processor 2 actuates the first switching unit 3 and a second switching unit 6 to select either a first 4 or a second acoustic analyzer 5. Each of the first 4 and the second acoustic analyzer 5 separates the speech data at equal intervals of substantially 10 ms on the basis of a Hamming window of about 25 ms and subjects its data segments to acoustic analysis such as cepstrum analysis to produce a train of acoustic vectors.

It is now noted that the speech may be free from or contain a noise in a particular frequency range of voice signal depending on the route of transmission or the country of the caller, for example, any call from a specific nation in Europe carries such a noise. For handling the former and the latter, the first 4 and the second acoustic analyzer 5 respectively are connected in parallel for selective use.

The first acoustic analyzer 4 analyzes the speech which contains no such noise. The second acoustic analyzer 5 has a notch filter or the like for removing the noise to produce a acoustic vector train from the noise-free speech. The embodiment is not limited to the two acoustic analyzers 4 and 5 shown and three or more acoustic analyzers may be used when three or more noise-imposed speech data are received.

- 5 -

There are provided three, first, second, and third, reference speech model storages 7 to 9 for saving speech models, e.g. HMMs, defined according to the countries of callers and the routes of transmission. A third switching unit 10 is responsive to a control signal from the line connection data processor 2 for selectively connecting one of the three reference speech model storages 7 to 9 to the speech pattern matcher 11. The speech pattern matcher 11 then compares a speech model d from the selected reference speech model storage with the acoustic vector train c transmitted through the second switching unit 6 for speech recognition and delivers its result.

In the embodiment, the most appropriate acoustic analyzer and speech model storage can be selected corresponding to the line connection data including the country of a caller and/or the route of transmission. This allows the speech recognition to conform to the characteristics of the telephone line and the noise, hence enhancing the performance of the speech recognition system. The speech models saved in the first to third speech model storages 7 to 9 may be increased in quality through learning.

A second embodiment of the present invention will now be described referring to Fig. 2. The second embodiment is substantially identical to the first embodiment but characterized by another set of first to third reference speech model storages 12 to 14 which are then explained in more detail.

In particular, the first 12 and the second speech model storage 13 save noise models. A speech may include a silence pause where no sound is made. It is essential for the speech recognition to correctly

discriminate the silence pause from the other speech period. However, an intrinsic noise is sometimes imposed on the silence pause when the call travels through a particular route. For identification of such intrinsic noises, their models are saved in the first 12 and the second speech model storage 13. The switching unit 10 is activated by a control signal from the line connection data processor 2 to selectively connect either the first 12 or the second speech model storage 13 to the speech pattern matcher 11. The third speech model storage 14 saves speech models for use regardless of the country of a caller and/or the route of transmission. More specifically, the speech models saved in the third speech model storage 14 are identical to those, e.g. HMMs, saved in a speech model storage 34 shown in Fig. 3. The speech pattern matcher 11 identifies the silence pause in a speech from the noise models supplied through the third switching unit 10 and collates the speech with the speech model from the third speech model storage 14 to recognize voice sounds in the speech. A result of the speech recognition is then delivered as an output.

The first 12 and the second speech model storage 13 can improve the noise models through learning. In practice, the speech data received from a telephone line is saved in memory means and when the telephone line is disconnected, is fed to the first 4 or the second acoustic analyzer 5 for extracting its noise data which is then saved. In this manner of learning, the noise models saved in the first 12 and the second speech model storage 13 can be improved.

While the speech model storage 14 saves the speech models for use regardless of the country of a caller and/or the route of transmission

- 7 -

in the previous embodiment and is of no limitation, a group of reference speech model storages may be provided for saving various speech models corresponding to the countries of callers and the routes so that one of them can be selected by a control signal from the line connection data processor 2.

The action of the second embodiment is explained. When a call is received from the international telephone line, the telephone line interface 1 identifies the calling country or city and/or the route of transmission from the telephone number of a caller. A resultant line connection data a is transmitted to the line connection data processor 2. Simultaneously, the telephone line interface 1 delivers a speech data b to the first switching unit 3. In response to the line connection data a received, the line connection data processor 2 sends selection control signals to the first 3, the second 6, and the third switching unit 10. For example, when the calling country or city is other than the specific ones, the first acoustic analyzer 4 and the first reference speech model storage 12 are selected by the control signals.

The speech data from the telephone line interface 1 is thus fed to its pertinent, first acoustic analyzer 4 where it is acoustically analyzed so that while silence segments in the speech data are detected by referring to the noise models saved in the first reference speech model storage 12, speech recognition is made from the speech models saved in the reference speech model storage 14.

If the call is received from the specific country or via the particular route, the second acoustic analyzer 5 and the second speech model storage 13 are selected. As the result, data of the call received

from the specific country or via the particular route can be subjected to the speech recognition with the most appropriate sections of the acoustic analyzer and the speech model, whereby the quality of the speech recognition of the data will be increased.

5

As set forth above, the embodiment allows the acoustic analyzer and the reference speech model storage to be selected according to the country of a caller and/or the route of transmission, thus enhancing the quality of speech recognition. When a not-continuous insertion signal is contained in a call from the public booth in a specific country, it will be eliminated by selecting and using desired one of the acoustic analyzers. In a domestic use, a call from a mobile telephone is identified by its key-station number and can be subjected to the speech recognition with the use of speech models for mobile telephone, thus having a higher level of speech recognition.

10

15

Embodiments of the present invention allow the line connection data in a call, including the telephone number of a caller and/or the route of transmission, to be used for selecting the acoustic analyzer and the speech models. Accordingly, the speech recognition is performed in reference with the line characteristics and the noise characteristics and will be increased in the quality.

20

CLAIM

1. A telephone speech recognition system for recognizing
speech data received from a telephone line comprising:
a telephone line interface connected to the telephone line for
detecting line connection data;
5 a plurality of acoustic analyzers having means for removing
noise derived from at least one of the line characteristics and/or the route
characteristics; and
a line connection data processor for selecting the acoustic
analyzers in response to the line connection data from the telephone line
10 interface.
2. A telephone speech recognition system according to claim 1,
further comprising:
a plurality of reference speech model storages for saving speech
15 models corresponding to at least one of the line characteristics and the route
characteristics, in which a selection signal from the line connection data
processor is used for selecting the reference speech model storages.
3. A telephone speech recognition system according to claim 2,
wherein the reference speech model storages save speech models
20 attributed to the telephone line.
4. A telephone speech recognition system according to claim 2,
wherein the reference speech model storages save noise models
appearing in silence segments in dialogue so that any silence segment in
a dialogue can be recognized by a speech pattern matcher referring to the
noise models.

5. A telephone speech recognition system according to claim 3 or 4, wherein the speech models saved in the reference speech model storages are updated by learning.

5 6. Telephone speech recognition system as hereinbefore described with reference to Figures 1 to 3 of the accompanying drawings.



The
Patent
Office

Application No: GB 9717929.5
Claims searched: All

Examiner: Al Strayton
Date of search: 6 November 1997

Patents Act 1977 Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:
UK CI (Ed.O): H4K:KBHE; KBHF; KBHX; KF42; KF50A,C,X; KOE
Int CI (Ed.6): H04M
Other: ONLINE: WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	GB 2 260 670 A (NORM PACIFIC)	
A	GB 2 194 412 A (HASHIMOTO)	
A	US 5 479 490 (NAKASHIMA)	
A	US 5 274 695 (GREEN)	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)